DOCUMENT RESUME

ED 163 048 TM 008 076

AUTHOR Bedard, Roger
TITLE Team Evaluating: An Approach to Item Sampling Implementation in Classroom Evaluation.
PUB DATE Jun 78
NOTE 17p.; Paper presented at the First Congress on Education of the Canadian School Trustees' Association (Toronto, Ontario, Canada, June 17-21, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; *Item Sampling; Learning Processes; *Student Testing; Summative Evaluation; *Test Construction; Test Items

ABSTRACT

This paper briefly discusses the use of item sampling for student evaluation in a summative evaluation context. Its proposition is that item sampling is a means of achieving better and broader student evaluation because it leaves room for individualized evaluation. Item sampling can be implemented properly by associating it to the concept of item domain. It is suggested that a team of three or more experienced teachers perform item sampling, and that further research is needed. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163048

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Canadian School Trustees'
Association's

First Congress on Education

Educational Policy and the choice of future
June 17th to 21th, 1978
Toronto, Canada

•

Team evaluating: an approach to item sampling implementation in classroom evaluation

Roger Bédard

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roger Bédard

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM.

TM008 076

Université du Québec
Télé-université

3108, chemin Sainte-Foy
Sainte-Foy, Québec
G1X 1P8

With the increased emphasis on individualization in the schools it may very well be that having all students responding to the same set of test items will soon become the exception rather than the rule!

Thomas R. Knapp (1973)

INTRODUCTION

This paper deals with item sampling application to student evaluation in a summative evaluation context. Its proposition is that item sampling is a mean of achieving better and broader student evaluation in the sense that it leaves room for individualized evaluation. It examines a way to implement item sampling properly by associating it to the concept of item domain. It concludes by suggesting that item sampling should be put to action-research work before extending its use.

GENERAL CONTEXT

The purpose of this paper is to framework offer one practical avenue to reach partially the student evaluation framework set by the United States National Education Association in 1975 and expressed in the following terms:

- *Frequent and comprehensive evaluation of student learning progress*
- *Evaluation of students mainly for the purposes of helping them to make wise choices and to learn better, and of helping teachers to better understand student learning problems and to prescribe for alleviation of those problems*
- *The use of a broad range of systems and procedures for evaluating student learning progress, including some types of tests*
- *The use of some individual diagnostic instruments, carefully developped*

criterion-referenced tests, and teacher-made tests. [1]

LIMITATIONS

Within this paper, we have limited our evaluation context to the student evaluation in a subject matter in which student performance can be extrapolated through his performance at a test. It is also assumed that teacher-made evaluation instruments are valid enough to provide accurate and dependable estimates of student performance. Due to the fact that little research is available on the subject, the type of evaluation (testing) suggested here is proposed first for experimentation or action-research. Finally, it encompasses only cognitive learning and leaves out affective and psycho-motor learning areas; evaluation in those areas should not be based solely on tests or exams of the paper-and-pencil type.

LEARNING PROCESS AND EVALUATION

The concept of learning process which evaluation is part of, is restricted to these situations where teaching is involved; this restriction necessitates that a third party, the teacher, is part of the action. The teacher is active at the three stages: first, he participates in the definition of the set of objectives, second, he develops the ordered series of learning tasks and, third, he verifies the attainment of the objectives by the students (see table 1). In

1) Report of the NEA Task Force on Testing. Reports of Committees, Councils, and Task Forces, 1974-75. Washington, D.C.: National Education Association, 1975. pp. 79-80.

schools today, the learning process is subsumed within an instructional program.

Table 1
Learning process when teaching is involved

Departure point	Journey	Arrival
A set of known objectives	Ordered series of learning tasks	Attainment of the objectives

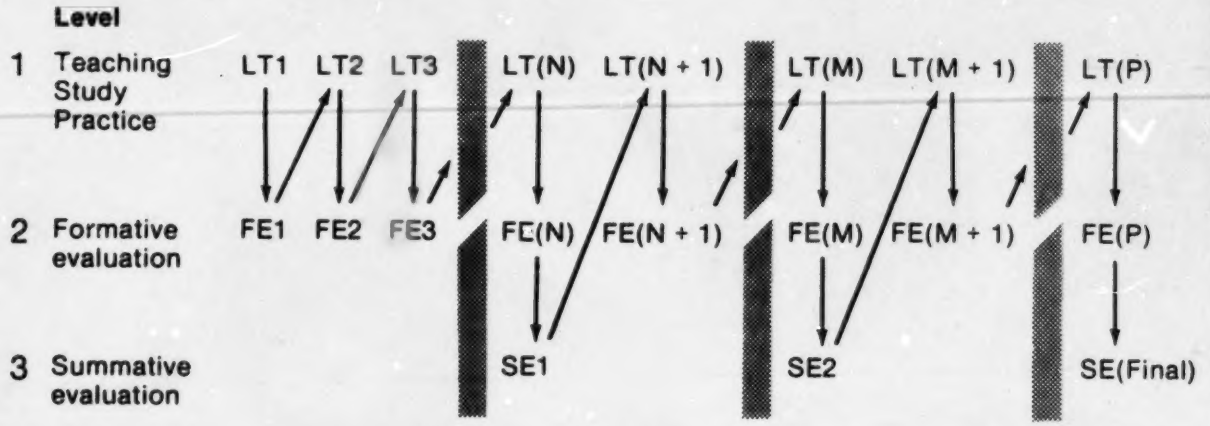
A learning task is seen as an organized number of activities (teaching, and/or study and/or practice) which are aiming at the acquisition by the students of the knowledge and abilities (see table 2).

Table 2
Learning task

Finite subset of objectives	▶	Teaching Study Practice	▶	Attainment of the objectives

Embedded in the learning process is the formative evaluation; its role is to help students *"to make wise choices and to learn better"* [1] and to inform teachers of learning problems students are faced with. Although not essential to the learning process, summative evaluation is still present in our educational system and it has to be performed either by the teacher himself or by external evaluators. Table 3 shows the connections between these three levels of intervention in the learning process. Summative evaluation is based on periodic exams or other appropriate measuring instruments.

Table 3
Instructional programs



1) Op. cit.

In this paper, we will suggest item sampling as a mean to achieve summative evaluation in a way that is better adjusted to the integrated learning process as proposed hereinbefore.

ITEM DOMAIN

It is a strong assumption of this presentation that item sampling should be employed in conjunction with an item domain. The evaluation framework adopted here is based on an approach proposed by Shoemaker (1974).

A teacher teaching a certain subject matter during one school year is working within an instructional program. Nothing new there, unless this instructional program is so closely related to evaluation as to be defined in relation with an item universe, *"a collection of stimuli for eliciting responses from examinees."* (Shoemaker, 1974) In fact, Shoemaker (1974) defines an instructional program as *"the vehicle for providing the necessary skills to answer correctly all items in the item universe."* (p. 2).

The item universe is not a workable concept. The item domain is a more manageable concept since it is a *subuniverse* containing a large number of items extracted from the universe. Of course, an item domain is assumed to be a representative sample of its parent universe. Shoemaker (1974, p. 5) suggests that both the instructional program and the item domain should be developed concurrently: first, determine the objectives of the instructional program, second, *"construct a given number of items for each objective"* and, thirdly, *"produce the instructional sequence and materials."*

When all the items constructed are catalogued and ready for

selection and use, they are referred to as an item bank. Item banking being only one approach to item domain, this last concept will be retained throughout this text.

An item domain can be subdivided into subdomains so that exams may be constructed on specific points of an instructional program. A specific series of exams is constructed by sampling items from the subdomain *"either through a random or stratified-random sampling plan."* (Shoemaker, 1974, p. 10) So, for the type of trial implementation of item sampling proposed here, it is possible to develop a subdomain of items on a limited part of the instructional program. Also, since the instructional program, in most cases, exists the teachers construct the subdomain straight off it.

ITEM SAMPLING

. *Definition*

Item sampling, more exactly named item-examine sampling, *"(...) applies to any measurement situation in which all of the individuals being tested are not administered the same set of test items (...)"* (Knapp, 1973, p. 1) Each *"set of test items"* becomes an exam. In an achievement testing situation, a teacher may prepare as many different exams as his class numbers students, but he would more likely construct two or three forms of his exam in a fashion suitable for the aims he has in mind.

. *Application to achievement testing*

Item sampling as a classroom evaluation technique was first applied by Poggio et Glasnapp (1973). They found that item sampling

"as a method for measuring classroom achievement provides more precise information [prediction value on the final exam] (...) than tests of the same length constructed in the traditional manner" (p. 5).

However, care must be exercised in using item sampling for classroom evaluation. In 1976, two studies (Rothrock and Michael, 1976) and (Forester and Michael, 1976) yielded results which show no real differences in prediction of success accuracy between constant-item test forms and item-sampled test forms.

In these three researches, the authors developed their experiment by applying item sampling directly to specific test construction. Prior to item sampling, there is a very important stage that should not be leapt and it is the identification and building of the item domain. Shoemaker (197?) goes so far as to write " (...) if achievement testing (...) is to advance significantly, it will do so only through coming to grips with item domains - *not* through the continual refinement of any small number of tests associated with an instructional programme." (p. 240) The proposal here is to come to item sampling via item domain. It is believed that is a more secure approach capable of better measurement objectives attainment and more permanent in terms of its ratio workload-length of usefulness.

. *Procedure*

Assuming there is a sufficient number of items, the application of item sampling is a rather straightforward operation. Three steps may be identified as such: the *sampling of items*, the *construction of the exams* and, finally, the *allocation of exams to examinee*. One avenue will be suggested hereafter not impliing its value compared with others.

The sampling of items is divided into two tasks: the first one consist in extracting the appropriate items from the item domain and, the second, is selecting the items for the exams. For a particular exam there exists a particular subset of items directly related to the portion of instructional program being examined; this subset must first be identified. Thereafter, a random sampling of items is performed in order to get the necessary number to construct the different exams.

Once the random selection of items is completed, one may proceed again by random sampling of these items to each exam; that is the first task in constructing the exams. Afterward, the teacher goes through the usual steps in order to get each exam in a shape ready to be administered; among others let's cite instruction writing, typing, proof-reading, duplicating.

Lastly, the different exams must be allocated randomly to examinees. The simplest way to realize this part is to pile them in a random fashion and distribute them as piled. A longer procedure, could be the random pairing of exams and examinees and the pre-addressing of each exam to its subgroup of identified examinees.

. *Advantages*

There are two main advantages to the implementation of item sampling in achievement evaluation; they are:

- Item sampling is a procedure "*for coming to grips with item domains*" (Shoemaker, 197?, p. 227)

- Item sampling allows for the evaluation of the instructional program.

If the ideal goal of achievement testing is to measure the performance of the examinee on the whole item domain associated with the instructional programs, the item sampling procedure offers a statistically unbiased way of estimating that performance, particularly because it allows for the use of more items and a more involved procedure of designing them.

It is also interesting to note that item sampling is one step toward individualized examination, that is each student is examined in accordance with the individualized instruction he got and his abilities.

The large number of items used in item sampling and the statistical procedures available [1] offer varied information useful in program evaluation. It is a helpful tool for program revision as long as the item domain is in this isomorphic relationship with its instructional program that Shoemaker (1974, p. 2) wrote about.

Finally, item sampling has statistical advantages: first, its usefulness for item analysis, and second, for a given number of observations the standard error of estimate for each paramater of an exam will be less.

1) Refer to Shoemaker, 1973,

. *Disadvantages*

In the context of achievement evaluation, item sampling when coupled with an item domain is a heavy burden for a teacher who already has a workload which leaves him little free time. That is the reason why it is thought that item sampling is best applied where three or more teachers can benefit from their common work.

Due to its logistical implications, item sampling can hardly be developed unless teachers can get clerical support from the school office of the principal. That again might preclude item sampling from smaller school where office staff is not available.

Conjunctural is the fact that item sampling as of now still has to be experimented in the context of item domain. This paper hypothesized that good results will come out; it is still to be proven.

TEAM EVALUATING BY ITEM SAMPLING: A TIME-STUDY

Lets create a simulated situation where three teachers wish to try evaluating their students using item sampling.

Lets suppose:

1- They all teach the same subject matter at the same level in equivalent classes in the same school for their third year

2- Having a total of 90 students, they wish to prepare six tests of 30 items each

3- The portion of subject matter they choose needs 15 class hours, 10 of these are used to present new material

4- They total 15 years of experience over which they have accumulated the equivalent of 100 items each

5- For typing, they may employ one of the school secretary.

First, they have to create their item domain (bank):

. Its size is so determined: 10 hours of presented material multiplied by 50 items per hour equal 500 items.

. They already have 300 items all together that they will will have to sort out and classify; it will take three hours each to perform this task

. They end up with 250 items retained

. They gather 75 more items through books, working as a team during four hours

. Each must write 25 items and spends three hours doing it

. So each teacher will have to find 10 hours so that a 500-item bank is created.

Over a period of three weeks, each teacher shall have to devote *3 hours and 20 minutes* per week to build a 500-item bank put at his disposition.

Second, they will construct their six exams:

- Supposing they want six equivalent exams, all they have to do is to select at random 180 items and assigned them randomly 30 to each test

- A three-hour meeting should allow sufficient time to perform this task

- Another three hours should be sufficient for instructions-writing, format selection, typing-instructions and proof-reading

- With two additional hours each, they are ready to enter the examination room.

Over a period of three weeks, each teacher shall have to devote *5 hours and 20 minutes* per week to build a 500-item bank and six exams constructes randomly from this bank.

Lets round off the figure to *6 hours* in order to care for forgotten factors and still we come only to less than 18% of the working time in order to possess the 500-item bank. This importance given to summative evaluation is nonetheless far too great but this simulation demonstrates the necessity of team work in order to implement satisfactorily summative evaluation by item sampling. The time estimate we have come to is showing some possibilities of two or three, at the most, trials of item sampling over an academic year

but not very much more. On the other hand, item sampling is gaining its full value when it can be mastered over the whole period of a course. As a conclusion, item sampling is of the size of a large school system project supervised by the specialist in evaluation.

CONCLUSION

A combination of item domain and item sampling is hypothesized to yield a better summative evaluation set up not precluding a better formative evaluation. Care should be exercised in implementing such a procedure particularly because of the large-scale work involved and the absence of concluding research results. Item sampling should be attacked by a team of teachers - at least three - and, preferably, with team-teaching experience.

BIBLIOGRAPHY

Forester, D.L. and W.B. Michael. A comparison of the Validities of a Traditional Item-Allocation Method and of an Item-Sampling Procedure in the Construction of In-Class Examinations in Intermediate Algebra. *Educational and Psychological Measurement*, 1976, 36, 933-38.

Knapp, T.R. Item Sampling. Unpublished manuscript, University of Rochester, Sept. 1973.

Osburn, H.G. Item Sampling for Achievement Testing. *Educational and Psychological Measurement*, 1968, 28, 95-104.

Poggio, J.P. and D.R. Glasnapp. Item-Sampling as a Classroom Evaluation Technique. Paper presented at the annual convention of the NCME, New Orleans, La., February 1973.

Quinto, Frances and B. McKenna. *Alternatives to Standardized Testing: A Handbook*. Washington, D.C.: National Education Association, 1977.

Rothrock, Julia E. and Michael, W.B. The Comparative Validities of Traditional Item-Selection and Item-Sampling Procedures in Evaluation of Student and Class Progress Toward Course Objectives in Freshman Psychology. *Educational and Psychological Measurement*, 1976, 36, 925-932.

Shoemaker, D.M. Toward a Framework for Achievement Testing. Rough draft of a technical report. California: Southwest Regional Laboratory for Educational Research and Development, 1974.

Shoemaker, D.M. Applicability of Item Banking and Matrix Sampling to Educational Assessment. In De Gruijter, Dato N.M. and L.T.Th. van der Kamp, *Advances in Psychological and Educational Measurement*. Toronto: John Wiley, 197?.

Shoemaker, D.M. *Principles and Procedures of Multiple Matrix Sampling*. Cambridge, Mass.: Ballinger Publishing Company, 1973.